VOLUME 63
NUMBER 10

WHOLE NO. 305
1949

Psychological Monographs:

General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology* with the *Psychological Monographs*

HERBERT S. CONRAD, *Editor*

The Development and Validation of a Set of Musical Ability Tests

ROBERT W. LUNDIN

Hamilton College
Clinton, N.Y.

Accepted for publication, June 5, 1949

Price $1.00

Published by
THE AMERICAN PSYCHOLOGICAL ASSOCIATION
1515 MASSACHUSETTS AVE., N.W., WASHINGTON 5, D.C.

TABLE OF CONTENTS

THE DEVELOPMENT AND VALIDATION OF A SET OF MUSICAL ABILITY TESTS

A. Introduction

THE best known of the so-called musical aptitude tests is the *Seashore Measures of Musical Talent* (28) (29). Since their publication in 1919, they have been given to many individuals and subjected to considerable research, so that at the present time, we are able to formulate an opinion of their value and limitations. Adequate descriptions appear in (8) (21) (24) (29). Reliability coefficients have been reported by (3) (6) (8) (12) (14) (20) (32). Saeveit, Lewis and Seachore (24) report reliabilities for the revised version which are considerably higher than those reported by investigations on the 1919 version (pitch .88, .78, loudness .88, .77, time .75, .70, timbre .74, .72, rhythm .62, .72, tonal memory .88, .89). Validation on the revised tests is based on internal criteria. Seachore insists that this is the only justifiable validation procedure. Other validation procedures have been carried out by (2) (3) (12) (18) (19) (20) (23) (25) (33) (37). Examination shows validity coefficients lower than those for reliability. Expect for tonal memory, the Seashore tests deal with sensory acuities and do not touch on such functions as interval discrimination, harmonic sequences, tonality or resolution. Thus, it would seem that the tests measure merely the ear's responsiveness to certain differences in the sound wave.

The *Kwalwasser-Dykema Music Tests* (13) in many ways are similar to the Seashore tests. Reports show them to have, in general, even lower reliabilities and validities than the Seashore (1) (4) (5) (9) (17) (26) (33) (34) (35). Other music tests have been devised by Schoen (27), Drake (5), and Madison (16).

1. *Theoretical Implications*

(*a*) *Theory of specifics:* The construction of the Seashore tests is based on the belief that musical ability, per se, consists of a number of sharply defined behaviors, or "talents" (as he calls them) (30). Of these talents, the Seashore tests purport to measure six. There are probably more which as yet remain unrecognized or unmeasured. These specific talents are relatively unrelated and may be present or absent in various degrees. The man who possesses most of these talents to a high degree should be the better musician. According to Seashore's theory, they are inherited qualities and hence are little affected by training.

(*b*) *Integrative theory:* Drake (6) subcribes closely to the Seashore theory of specific talents, adding, however, that they are all dependent upon or knit together by the factor of musical memory.

(*c*) *Omnibus:* Mursell (22) opposes a thesis of specifics, putting forth a view which Seashore has labeled an omnibus theory. Taking a Gestaltist viewpoint, Mursell believes that musical behavior is dependent on the total energies of the individual and not on some special talents. So that for Mursell, musical ability in a general sense consists of a number of interrelated behaviors. From this viewpoint he has attacked the low validities of the Seashore tests (22). Seashore validated his tests on internal item consistency claiming that the tests do purport to measure the specific talents named as ability to discriminate differ-

ences in pitch, intensity, time, etc. This is undoubtedly true, but if they do not relate to our ultimate goal as best validated on external criteria, of what value are they as music ability measurements?

When we come to Seashore's premises, we realize why his tests have failed as valid instruments of measurement in music. There is no doubt that the tests measure five kinds of sensory discrimination, but the validation studies have shown that with the exception of pitch and tonal memory these tests are not related to the necessary components of musical behavior. Mursell has criticized the Seashore tests severely but has designed no substitute measurements. We agree with him to the extent that musical behaviors are interrelated to a great extent. He insists, however, on the partial inheritance of musical ability. Just what is inherited has not yet been shown.

(*d*) A different approach to the problem seems advisable in view of previous failures. If we consider that musical behavior is acquired through a long process of individual interactions with musical stimuli, the type of measure we choose to construct in order to test some of the resulting behaviors, should be of a different nature than most of those set down by Seashore and Kwalwasser. We need not accept the inheritance of capacity for music further than realizing the necessity for sound biological equipment for the reception of stimuli and performance at an instrument. The man born deaf is deprived of part of his biological equipment with which he may acquire musical behavior. Although Beethoven wrote some of his greatest work during his later years while deaf, we must consider that his equipment had already been acquired and developed before he became so stricken.

We should not be led to believe that musical ability is the result merely of classroom achievement. We realize that before the onset of formal training individual predispositions toward music will vary. Family stimulations early in life should not be ignored. Consider such great masters as Bach or Mozart. Both came from musical surroundings and stimulation came early in life. By the time a music instructor is sought, a wide variety of individual differences in musical behavior-equipment has been established. Musical ability, then, is not a single capacity possessed in various degrees by individuals. It may consist of a number of acquired behaviors built up through a process of interaction of the individual with musical stimuli over a period of time. In any attempt to measure such behavior we merely attempt to select some possible behaviors for our consideration. Any set of measures, therefore, can hardly tell the whole story. Because there are many kinds of musicians as the performer (violinist, pianist, or vocalist), the composer, theoretician or musicologist, the behavior equipment necessary for achievement will differ. Even among the performers, behaviors necessary for success will not always be the same. Ability to discriminate fine differences in pitch will be a prerequisite for the successful violinist or vocalist but not necessarily important for the pianist. Nevertheless, we believe there are some behaviors which musicians share. These are taught almost universally in music schools and usually are listed under the title of *theory*. The degree to which one will acquire such behavior will depend, however, on his previous equipment. Examples of such behaviors are writing melodies and

harmonies correctly after they have been produced audibly, harmonizing single melodic lines correctly following the rules set down by the older masters, ability to play and write rhythms correctly, and ability to detect changes in sequential patterns.

B. Purpose

This investigation had as its primary aim the construction and validation of a group of tests which would measure directly and in an objective fashion some of the kinds of musical behavior not heretofore considered by previous investigators, and which we believe are important constituents of a musical personality. These abilities have been characterized above. Secondly, we wished to consider the relationship of these musical behaviors to intellectual behavior and such sensory acuities as measured by Seashore and other investigators. Finally, we hoped that these tests, when used separately or along with other satisfactory measures, would serve as a guide in the selection of music students. A set of tests which are designed to measure some musical behavior in its actual setting should, if properly constructed, give us a more valid indication of what constitutes musical ability than those previously attempted.

From the literature we have noted that few tests measure musical ability in its actual setting. We noted that most of the measures other than sensory discriminations depend on subjective criteria, such as musicians' opinions in determining the correctness of an item (for example, the Kwalwasser tests of tonal movement and melodic taste, the Drake test of intuition, and the Schoen test of tonal sequence).

C. Description of the Tests

The five tests to be described are an attempt to measure musical behavior in a preliminary research undertaken in 1943 (15).

In each test the subjects were given three practice trials before starting on the test proper. The number of each item was announced on the phonograph record before the subject heard the music which he was to judge. This, as in the Kwalwasser-Dykema tests, reduces the subject's chances of losing his place. In all cases the subject made a response of "same" or "different." The items were arranged in random order so that a subject would get no more correct than he deserved after the chance factor was allowed for.

1. *Interval Discrimination*

The first test, which we called interval discrimination, consisted in its final form of fifty pairs of tones or musical step intervals, each interval contained two notes. For each item there were two sets of step intervals. *S* was asked to tell whether or not the second interval was the same as or different from the first. When an interval was the same, the number of steps or notes on the scale which lay between the first and second note of each interval also was the same. However, this did not mean that the actual notes played were the same for they were not, but in some cases the intervals between the notes were. This test was divided into two parts (Ia & Ib). Half of the items were played in an upward progression, called ascending intervals, while the other half was in a downward progression called descending intervals. In an upward progressing interval, the second note always had a

pitch which was higher than the first. In the downward progressing intervals, the pitch of the second note was lower than that of the first. In the preliminary investigation (15) this test consisted of twenty-five items recorded on the piano with no differentiation being made between upward and downward progressing intervals.

2. *Melodic Transposition*

The second test, melodic transposition, consisted in its final form of thirty pairs of simple melodies. The second melody was always in a different key from the first. Sometimes the transposed melody was not the same as the first, there being changes of one or more notes, so that if the second melody had been transposed back in to the original key, it would have been different. If *S* considered the second melody the same as the first (except for the change in key), he would respond by an "s," but if he detected a change in the transposed melody, he would mark "d."

3. *Mode Discrimination*

Test three, mode discrimination, was a completely new test and replaced that test called harmonic transposition in the first study (15). This latter test was eliminated on the basis of its low reliability and difficulty. The test of mode discrimination consisted of thirty pairs of single chords. If both chords in any item were of the same harmonic structure, in the same mode, for example both major or minor chords, *S* would respond with "s," or if one chord was major and the second minor, he would write "d." The test was similar to number two to the extent that single chords were transposed instead of simple melodic lines. If the transposed chord was the same except for change in key, it was an "s" item, and if the chords were different in harmonic structure, it was a "d" item.

4. *Melodic Sequences*

The fourth test was called melodic sequences. *Groves' Dictionary of Music and Musicians* (10) defines a sequence as ". . . the repetition of a definite group of notes or chords in different patterns of the scale like regular steps ascending or descending." This test consisted of thirty sequential groups or patterns. Each item contained four such groups. In all cases the first three patterns followed the same melodic order, but sometimes the last group did not follow the same pattern as the first three. In such cases the subject responded with a "d." Thus, if the entire sequence seemed correct, he replied with "s." All sequences were diatonic, that is, "in key" and began and ended in the key of C. Here we speak of diatonic as opposed to modulating sequences which change key with each repetition of the pattern. It was impossible to use modulating sequences when only melodies are played and still keep an objective criterion of correctness.

5. *Rhythmic Sequences*

Test five, rhythmic sequences, was completely new and replaced the test of harmonic sequences in the original study. This latter test was the poorest of the original battery from the point of view of reliability, difficulty in selecting items, and lack of discriminative value of items used. In this new test of sequential rhythm, four patterns were played as in the previous test of melodic sequences. The rhythmic pattern was set by the first three sequences, and *S* was asked to judge whether or not the last rhythmic

group followed the same pattern as the first three. This measure is different from any previous rhythm test in that it does not isolate the rhythm from the melody. Since rhythm in music seldom occurs in isolated patterns, this kind of measurement seemed a better index to use.

D. Test Procedure

The five tests constructed were decided upon after consultation with members of the Indiana University Department of Psychology and School of Music. (Items for the preliminary tests were composed by the writer. These preliminary tests should not be confused with the study done in 1944. This previous investigation served merely as a guide to the reconstruction of items.) The test items were played before a committee of three theory professors from the School of Music, who graded them as to difficulty and discriminative value. An equal number of items was made at each level of difficulty; very easy, easy, medium, difficult and very difficult, as judged by the instructors. Those items which were believed extremely easy or difficult, confusing or incorrect, were thrown out.

The 1944 study (15) was valuable in pointing out certain deficiencies and limitations which were kept in mind in the construction of items for this revision. First of all, we had found we were not measuring over a wide enough range of talent, and an item analysis indicated certain items to be more discriminative than others.

The final selection of preliminary items consisted of 100 for test I (50 ascending and 50 descending), 50 for test II, 80 for test III, 59 for test IV and 60 for test V. These were tentatively arranged according to difficulty as estimated by the committee and the writer. Tests I-IV were recorded on twelve-inch phonograph records by the American Recording Studio, Indianapolis, Indiana, using a Hammond electric organ as the medium for sound. The simple diaphasian organ stop provided the quality of tone. Test V was recorded in the Indiana University Radio Studio using a piano. The piano was selected for the rhythm test because the accenting of tones, a factor essential to rhythm, is more easily achieved on this instrument. The organ was selected for the first four tests because it is an instrument which controls volume regardless of the force with which a key is struck by the artist. This controlled intensity was a variable particularly important in test III, where four notes were played simultaneously, A time interval of three seconds was allowed between items. This was regulated by use of a stop watch. We had found this to be an optimum time interval after preliminary investigations. Following the recording process, the tests were given to two trial groups. Group I consisted of 60 students from two elementary laboratory courses in psychology. The second group contained 15 musicians from the School of Music selected by their instructors because of good work. On account of the length of the preliminary tests, they were administered in two sessions of about one hour each.

1. *Item Analysis*

Items were analyzed for each test in a twofold manner: first, for internal consistency with the rest of the subject, and second for degree of difficulty. In order to get a measure of internal consistency, the Guilford Phi Coefficient (11) was used. These correlations were computed for group 1 only. The proportion passing

each item was computed for each group separately. This was done to give an indication of the differences between groups and to determine the upper and lower limits of talent.

Of the remaining items, we then selected those which gave (*a*) the highest correlation with the rest of its sub-test and (*b*) which showed the greatest difference between groups in proportion passing. By this twofold criteria we hoped to obtain items which were not only internally consistent but discriminative. In almost every case we found that items which were the most discriminating also gave the highest correlation with the rest of the test.

After final selection of items, the tests were again recorded, duplicating previous conditions as much as possible. We used the same recording artists, recording studios, organ stop and time interval between items. The tests in their final form consisted of five ten-inch and one twelve inch records. Answer sheets were provided on which subjects were presented item numbers for each test with the symbols S or D after each item. Instructions required the subjects to encircle with pencil either letter according to his judgment for that item.

2. *Subjects and Administration of Tests*

For purposes of statistical analysis, two groups of subjects were selected. Group I, whom we shall call musicians, consisted of 167 full time students from the Indiana University School of Music. These students intend to follow professional careers either as performers or as teachers. The group contained about 90% of the present music school enrollment. Group II we shall call unselected, because they were chosen without regard to musical training. This group was made up of freshmen taken from one of the elementary psychology classes at Indiana University. They were 196 in number.

The tests were administered to the musicians during class time with the cooperation of instructors. For the musicians, the test administration including instructions and playing of records takes approximately fifty minutes for any one group.

In selecting a typical group of college freshmen without regard to musical training we chose a freshman class in psychology because it draws from all schools of the University. The students in this group took the tests outside of class in one of three meetings. Tests were not given during class time because the group was too large, and because of the extra time needed to instruct people unfamiliar with music. We found from experience in the preliminary trial that the tests could not conveniently be given in a fifty-minute class period. Students were given class credit for attendance at the test period. (Attendance was not on a voluntary basis, neither was it compulsory without some compensation. We felt the selection by mere volunteers would load our group with people who might have both interest and training in music. The approximate administration time for this group was sixty minutes. The same instructions were read to both groups. Usually musicians clearly understood the instructions on first reading. However, with the unselected group, it was often necessary to supplement the verbal instructions with further comments and blackboard diagrams, particularly in the case of tests I and III. The three practice exercises given before each test served as a useful tool for subjects in understanding the procedure

followed in the test. Students, particularly in the unselected group, were encouraged to ask questions before or after practice exercises were played. On their answer sheets students were asked to designate their class standing so that we could eliminate from the unselected group any students who were not freshmen. We desired a group homogeneous in this respect on which to standardize our results.

3. *Other Data*

In both groups we obtained an indication of the liking toward classical music. We asked students to indicate their liking toward classical music according to the following scale: *dislike very much, dislike, indifferent, like, like very much.* The two groups were also asked to indicate the number of years of instrumental or vocal instruction which they had previously had up to the time of taking the tests.

E. Discussion of Results

1. *Reliability*

Split-half reliability coefficients for the separate sub-tests and total scores appear in Table I, computed for the musician and unselected groups separately. For the unselected freshmen, coefficients are .85 for total scores and .70 or above for separate tests, except in the case of Test III, Mode Discrimination. For the person completely naive concerning things musical, Test III apparently was difficult to comprehend. Therefore, it should be used discreetly for groups completely unfamiliar with music. The mean score for this test was 16.56 out of 30 possible correct responses. Looking at the reliabilities for the musician group we do not find a drop in the reliability of Test III as compared with the rest of the battery. The reliabilities in the musician group are slightly lower than those found for the unselected people. This is probably because here we have a narrower range of talent. However, when we take the *total* scores we find a *higher* reliability (.89) than that found for the unselected group—reflecting the greatly superior reliability of Test III in the musician group.

TABLE I

Reliability Coefficients for Music Tests, Computed by the Split-Half Method for Musicians and Unselected Groups Separately

Test	Musicians	Unselected
I. Interval Discrimination	.79	.71
II. Melodic Transposition	.65	.72
III. Mode Discrimination	.65	.10
IV. Melodic Sequences	.70	.77
V. Rhythmic Sequences	.60	.72
Total scores	.89	.85
	$N=167$	$N=196$

If we compare these results with the reliabilities reported for the Seashore and Kwalwasser-Dykema tests, we find they are superior to the reliabilities for the Kwalwasser tests in general, and compare favorably with reports on Seashore tests.

In general, our results indicate a battery of music tests which is reliable for general predictive purposes, particularly when total scores are used. We suggest that Test III may be omitted if the subject is completely unfamiliar with musical terms. However, if he has a basic knowledge of musical modes, that is, if he can distinguish a major from a minor chord, the test can be used to advantage.

2. *Validity*

The tests were validated on the musician group alone, since no adequate criteria were available for the unselected group. Teachers were asked to rate their students on a graphic rating scale devised by the author for this purpose. These ratings by professors were used as the criteria in validating the tests. Students were rated on the following musical behaviors; (*a*) melodic dictation, (*b*) harmonic dictation, (*c*) written harmonization, (*d*) general ability in theory, (*e*) vocal or instrumental performance. A total of the first four ratings constituted the sixth category (*f*) of total ratings. The ratings on performance could not be included in this total since data were incomplete ($N = 62$). In rating a student on melodic dictation the accuracy with which he was able to write melodies from dictation was the main point of consideration. This holds true for harmonic dictation except that here the student wrote chords from dictation. Ratings in written harmonization were based on how accurately the student followed the proper rules of harmonization and how musical were his results. The general ability rating was based on the students' grades and daily exercises. Performance was based on students' proficiency in performance of vocal or instrumental music whichever the case might be.

The raters were instructed to place a check mark anywhere along a line nearest to the description below the line which best fitted the student they were rating. For purposes of correlation, the distance from the beginning of the line to the point where the check was made, was measured with a millimeter ruler. This distance was the score. A copy of this rating scale appears in Figure I.

When students took the present tests, they were asked to indicate their theory instructors. If they had had more than one instructor in theory, they indicated as such. Whenever possible, more than one instructor rated the student and an average rating was taken. Performance ratings were made by instructors who were best acquainted with the students' proficiency. Each test (I*a*, I*b*, I, II, III, IV, V, Tot. Sc.) was correlated with each rating (Melodic Dictation, Harmonic Dictation, Written Harmonization, General Ability, Performance, and Sum of ratings). These results appear in Table II. The total test scores correlated highest with ratings on Melodic Dictation .70, Harmonic Dictation .70, General Ability in Theory .65, Performance .51 and Total ratings .69.

Of the individual tests, Test I (Interval Discrimination) correlates highest with ratings on Melodic Dictation (.66) and next with ratings on Harmonic Dictation (.60). This is what we should expect, because it seems reasonable that a student who can keenly distinguish between intervals of various sizes would be better equipped to write a melody or harmony on paper after hearing it played. Writing a melody from dictation involves discriminating between large and small intervals.

Test II (Melodic Transposition) correlates best with harmonic dictation (.52), Test III (Mode Discrimination) correlates equally well with melodic dictation (.51) and harmonic dictation (.51), Test IV (Melodic Sequences) is best related to melodic dictation (.57), harmonic dictation (.56) and to general ability in theory (.56) and Test V (Rhythmic Sequences) to general ability (.33).

FIGURE I

RATING SCALE FOR MUSICIAN

1. Melodic dictation: Consider the accuracy with which the student writes melodies from dictation.				
Very inaccurate	Occasionally gets dictation correct. Usually has a number of mistakes.	Takes a dictated melody with a fair degree of accuracy.	Takes dictation well; occasionally makes mistakes.	Writes melodies from dictation rapidly and without errors.
2. Harmonic dictation: Consider the proficiency with which the student writes harmonies from dictation.				
Seldom can recognize or write harmonies correctly.	Often recognizes harmonies, but makes errors when writing, (position chord).	Sometimes gets dictation right, other times is wrong.	Recognizes harmonies, but occasionally makes mistakes in writing them down.	Recognizes harmonies easily and writes them correctly.
3. Written harmonization: Consider how accurately he follows the rules of proper harmonization and how musical are his results.				
Inaccurate in harmonizations, has little feeling for what what would sound well.	Uses the standard run of the mill harmonies; frequently makes mistakes.	Generally follows the rules; harmonizations average in music value.	Harmonizations sound well, occasionally makes mechanical mistakes in writing them.	Follows all rules and makes clever harmonizations of melodies.
4. General ability in theory course: Consider his all around ability in course work as measured by tests and daily exercises.				
Does sub-standard work; is consistently poor.	Does passing work of a sub-standard quality.	Does average work on tests and daily exercises.	Work is usually good.	Does excellent work, is consistently good on tests and daily exercises.
5. General performance ability in instrumental or vocal music: Consider the degree of proficiency with which he plays or sings as the case may be.				
Has little or no aptitude for the instrument.	Has some ability but is below average.	Plays with an average degree of ability.	Performance on the instrument is good.	Has a great deal of facility in playing, is definitely superior.

TABLE II

VALIDITY COEFFICIENTS COMPUTED FOR SEPARATE TESTS WITH SEPARATE RATINGS, TOTAL TESTS, AND TOTAL RATINGS. ($N=167$)

	Melodic Dic.	Harmonic Dic.	Written Harmon.	Gen. Ability	Performance*	Total Ratings
Ia. Interval Discrimination (ascending)	.51	.56	.29	.45	.26	.68
Ib. Interval Discrimination (descending)	.51	.47	.27	.49	.21	.56
I. Interval Discrimination (total)	.66	.60	.32	.55	.45	.48
II. Melodic Transposition	.49	.52	.26	.44	.26	.45
III. Mode Discrimination	.51	.51	.35	.42	.35	.49
IV. Melodic Sequences	.57	.56	.45	.56	.38	.57
V. Rhythmic Sequences	.26	.26	.10	.33	.17	.26
Total Scores	.70	.70	.43	.65	.51	.69

* Data incomplete ($N=62$).

Test V correlates lowest with the criteria used. However, we should not necessarily interpret this to mean that a measure of rhythm is not an important test to be considered in measuring musical behavior. Perhaps all we should say is that we were unable to get an adequate criterion against which to validate this test. In setting up a rating scale we inquired of instructors concerning their ability to rate rhythmic behavior. In most instances their answer was negative. Therefore, we felt it inadvisable to include a criterion which we knew beforehand might be inadequate and therefore inaccurate.

In general, the validity of these tests is high, especially for the total scores. Our results are superior to those previously reported on the Seashore or Kwalwasser tests when external criteria were employed.

The ratings for written harmonization correlate, in general, lower with separate tests than other ratings. It may be that as yet we have not devised a test which measures as adequately as we should like this particular form or forms of musical behavior.

3. *Weighting the Tests*

Using as our criterion the sum of the ratings, we set up a Doolittle work sheet (Cf. Peters, C. C. & Van Voorhis, W. R. *Statistical procedures and their mathematical bases,* N.Y.: McGraw-Hill, 1940, p. 226 f.) for solving a multiple correlation problem. As a result of this procedure we set up a multiple regression equation in terms of raw scores for the various tests. When we solve for our multiple $R_{1.23456}$, we get a coefficient of .71, which is only slightly higher than that found when the tests are merely added numerically and their sums correlated with total ratings (.69). Therefore, we conclude that weighting the various tests adds little to their predictive value.

4. *Relationship between Test Scores, Training, and Interest*

On their test blanks, students in both groups were asked to indicate the number of years they had taken instruction in the performance of vocal or instrumental music. If more than one instrument had been studied, both were indicated and the years totaled. If the subjects had had no training, they indicated that by writing a "o" in the appropriate place. Then, the number of years training was correlated with total tests scores for each group separately. For musicians the correlation between these two variables is .43, for the unselected group, .38. We thus have a positive relationship which is not high.

Students were also asked to indicate their liking for classical music by making a check on a five point scale as follows:

Indicate your liking for classical music

dislike very much	dislike	indifferent	like	like very much

Converting to numerical scores of 1-5, we then correlated liking for classical music with total test scores. Correlations again are positive but low, and slightly lower than those for training. For musicians, r was .30, for the unselected group, r was .23.

5. *Differences between Groups*

The object of giving a test of this nature to a group of unselected subjects was first, to establish norms for a general population of college freshman against which any single score could be compared (see Appendix); and second, to determine whether or not there were statistically significant differences between the performance of musicians and unselected subjects on the various tests. The results appear in Table III. This table gives the means, standard devia-

TABLE III

MEANS, STANDARD DEVIATIONS, AND CRITICAL RATIOS FOR MUSICIANS AND UNSELECTED GROUPS

Test	Group	Mean	Standard Deviation	Critical Ratio
I*a*	Musicians	22.10	2.15	6.40
	Unselected	17.14	3.11	
I*b*	Musicians	21.36	2.75	7.70
	Unselected	16.17	3.15	
I*t*	Musicians	43.46	3.92	14.70
	Unselected	31.19	5.33	
II	Musicians	27.72	2.25	6.73
	Unselected	23.41	3.40	
III	Musicians	22.28	3.97	8.17
	Unselected	16.56	2.71	
IV	Musicians	26.79	2.56	7.90
	Unselected	21.26	4.07	
V	Musicians	27.19	2.68	5.09
	Unselected	23.06	6.04	
Total score	Musicians	147.68	11.02	21.51
	Unselected	118.43	14.17	

TABLE IV

INTERCORRELATIONS OF TESTS FOR UNSELECTED GROUP ($N=195$)

	Ia	Ib	It	II	III	IV	V
Ia		.92	.81	.41	.29	.46	.18
Ib	.92		.78	.44	.29	.43	.22
It	.81	.78		.48	.30	.47	.23
II	.41	.44	.48		.22	.63	.32
III	.29	.29	.30	.22		.52	.19
IV	.46	.43	.47	.53	.52		.31
V	.18	.22	.23	.32	.19	.31	

tions, and critical ratios for the two groups. We note that the table shows very significant differences between means on all five tests and total scores. In considering these differences, we should recall that the unselected group was drawn in such a manner that it could include some musically trained people. This is evidence that we are measuring behavior more typically found in a population of musically trained persons.

6. *Intercorrelations of Tests*

To find out whether or not each test was measuring the same or different kinds of musical behavior, the five tests were intercorrelated, (including I*a* and I*b*) for the musicians and the unselected groups separately. Tables IV and V show these results.

From our inspection of these tables we may observe the following trends:

1. All intercorrelations are positive and rather closely related.

2. The tests tend to intercorrelate slightly higher for the musicians than for the unselected group.

3. Tests I, II, and IV intercorrelate highest with each other for both groups.

4. Test V correlates the lowest with the other tests.

5. Tests I*a* and I*b* are highly related for both groups and as we should expect, are highly related to I*t*.

From these tables of intercorrelations we may say that while we may be measuring for the most part different kinds of musical behaviors, they are quite highly related, rhythmic sequences showing a lower degree of relationship to the others.

7. *Relation with Seashore and Drake Tests*

Data on the revised Seashore tests of pitch, rhythm and tonal memory and the Drake test of tonal memory were available for the freshmen and sophomores in the music group.

Table VI gives the correlations between these tests and our battery. The Seashore tests correlate low but positively with our tests. The Seashore tonal

TABLE V

INTERCORRELATION OF TESTS FOR MUSICIAN GROUP ($N=167$)

	Ia	Ib	It	II	III	IV	V
Ia		.68	.84	.46	.47	.50	.24
Ib	.68		.75	.37	.35	.50	.30
It	.84	.75		.48	.55	.39	.38
II	.46	.37	.48		.36	.53	.35
III	.47	.35	.55	.36		.49	.28
IV	.50	.50	.59	.53	.49		.39
V	.24	.30	.38	.35	.28	.39	

TABLE VI

CORRELATIONS OF SEASHORE AND DRAKE TESTS WITH LUNDIN TESTS

	Seashore Pitch	Seashore Rhythm	Seashore Tonal Memory	Drake Tona Memory
I. Interval Discrimination	.21	.20	.17	.55
II. Melodic Transposition	.19	.04	.20	.26
III. Mode Discrimination	.26	.26	.30	.56
IV. Melodic Sequences	.20	.34	.31	.47
V. Rhythmic Sequences	.22	.20	.29	.20
Total Scores	.30	.07	.31	.60
No. of Cases	105	104	100	104

memory test shows a slightly higher relationship with ours than pitch and rhythm. This we might expect since tonal memory is involved to some degree in our tests. The Drake test correlates higher particularly with I, II and total scores.

As a measure of the validity of the Seashore and Drake tests, we correlated these data with our ratings. We thus have a fair measure of comparison between the validity of our tests as compared with Seashore's and Drake's. This data appears in Table VII. If we compare this data with the validity coefficients of our tests (refer back to Table II), we notice that almost without exception for any one criterion, our tests have much higher validity coefficients, particularly for total scores.

Although the Drake tests do not have as high validity coefficients as ours do, they are superior to the Seashore tests.

8. *Relation with Intelligence*

Scores on the *California Mental Maturity* test were secured for a large proportion of both of our groups. Data on these tests were given in raw score form. This test gives scores for the Non-Language, Language, and Totals. Correlations between each of our tests and each part of the California test were computed for the musicians and unselected groups separately. These appear in Tables VIII and IX.

TABLE VII

VALIDATION OF SEASHORE TESTS OF PITCH, RHYTHM, AND TONAL MEMORY AND DRAKE TEST OF TONAL MEMORY

Test	Melodic Dictation	Harmonic Dictation	Criteria Written Harmonization	General Ability	Performance	Total Rating	N
Seashore Pitch	.32	.32	.28	.23	.13	.24	105
Seashore Rhythm	.35	.40	.19	.28	.35	.30	104
Seashore Ton. Mem.	.31	.31	.23	.32	.45	.31	100
Drake Ton. Mem.	.50	.45	.36	.42	.09	.47	104

We note for both groups that the relationships are low with no correlation over .25. We can make no statement concerning a greater or less relationship between our tests and the language or non-language part of the California test. The relationships differ from test to test and from group to group. For example, for the musician group with Test I, the correlation is slightly higher with non-language (.20) than language (.11). But

TABLE VIII

CORRELATIONS BETWEEN CALIFORNIA MENTAL MATURITY TEST AND LUNDIN TESTS FOR MUSICIANS ($N=113$)

	Non-language	Language	Total
	(California Mental Maturity Test)		
I. Interval Discrimination	.22	.11	.15
II. Melodic Transposition	.20	.21	.16
III. Mode Discrimination	.17	.04	.23
IV. Melodic Sequences	.20	.16	.25
V. Rhythmic Sequences	.10	.03	.03
Total Scores	.25	.13	.15

for the unselected group the reverse holds, non-language (.11) and langauge (.17). We must say, therefore, that general intelligence, as measured by the California test, shows little or no relationship with the musical behavior measured in the present battery of tests. This leads to the substantiation of the previous belief that little relationship exists between intelligence and musical ability. These previous findings were based, however, on data gathered from the Seashore tests, which we have seen are not valid when correlated against a criteria which we feel gives opportunity to rate a good sample of various kinds of musical behavior.

TABLE IX

CORRELATIONS BETWEEN CALIFORNIA MENTAL MATURITY TEST AND LUNDIN TESTS FOR UNSELECTED SUBJECTS ($N=155$)

	California Mental Maturity Test		
	Non-language	Language	Total
I. Interval Discrimination	.10	.17	.13
II. Melodic Transposition	.23	.19	.24
III. Mode Discrimination	.04	.03	.05
IV. Melodic Sequences	.16	.17	.22
V. Rhythmic Sequences	.22	.11	.18
Total Scores	.22	.19	.24

F. SUMMARY AND CONCLUSIONS

The present investigation had as its aim (*a*) the construction of a series of tests of musical behavior which would measure in an objective fashion some kinds of musical behavior not already considered by previous investigators; (*b*) the determination of the relationship between these tests and other existing music tests such as those by Seashore (24) and Drake (7); and finally (*c*) the determination of the relationship between the new tests and general intelligence.

The tests decided upon for construction bear the titles: *interval discrimination, melodic transposition, mode discrimination, melodic sequences,* and *rhythmic sequences.* For the preliminary test, items were selected by a group of three musicians and the author, and the tests were recorded on phonograph records, using a Hammond organ and piano as media for sound. This preliminary test was given to two groups (60 elementary laboratory students and 15

musicians) for purposes of item analysis. Items were selected for the final test which showed the greatest internal consistency and the greatest difference between groups. These items were then arranged in order of difficulty and re-recorded, duplicating as much as possible the conditions of the preliminary recording. The final tests contained approximately one-half the number of items as the preliminary tests.

The final tests were given to 167 students selected from the School of Music and 196 unselected freshmen from one of the elementary psychology classes. The unselected freshmen were chosen without regard to musical training.

Tests were validated against a criterion of five different ratings made by professors for the music group alone. Ratings were obtained for the music students' abilities in melodic dictation, harmonic dictation, written harmonization, general theory, and instrumental or vocal performance. The *sum* of the *first four* ratings constituted a sixth category.

Reliability coefficients were computed for each group of subjects separately. For each group, also, total scores were correlated against training and liking toward classical music. The correlations between the new tests and the Seashore and Drake tests and with the California Mental Test of general intelligence, were also determined. For comparison of validities, the Seashore and Drake tests were validated against the same criteria as our tests.

The results indicate reliability coefficients (computed by the split-half method for each group separately) that are high enough to be used for general predictive purposes, particularly when total scores are used. The reliabilities of the tests are superior to those found by previous investigators for the Seashore and Kwalwasser music tests. It is recommended, however, that for the individual completely naive to music that Test III (Mode Discrimination) be used with discretion.

Individual tests correlated highly in general with the criteria. For total scores, the validity coefficients are .70 for melodic dictation, .70 for harmonic dictation, .65 for general ability in theory. .51 for performance and .60 for total ratings. These coefficients are superior to those reported by previous investigators for other tests, when similar external criteria are used. This finding leaves little doubt that our tests are measuring more directly and accurately such musical behavior which is deemed important by instructors of music.

When tests are weighted using a multiple regression equation, results show such slight increase in predictive value that a mere summation of tests into a total score is equally satisfactory. The correlation between the sum of the tests and the sum of criterion-ratings was .69, whereas the multiple correlation between the criterion and the individual tests was found to be .71.

For both groups there is a positive relationship between total scores and number of years of training (for the musicians, .43; for the unselected group, .38).

There is also in both groups, a low but positive correlation between liking for classical music and total test scores, the relation being lower for the unselected than for the musician group (.23 vs. .30).

All the tests are positively intercorrelated. This holds true for both groups (See Tables IV and V.)

There is a statistically very significant

difference between the means of both groups for each test and for total scores. From this and other findings we may conclude that our tests are measuring behavior more typical of a group of musicians than of a population of unselected freshmen.

The relationship between the present tests and the Seashore tests of pitch, rhythm and tonal memory is low (tonal memory slightly higher than the others). There is also a slightly higher relationship between the Drake test of tonal memory and the present tests, than between the Seashore and the present tests. This slightly higher correlation may be accounted for by the fact that in answering items in the present set of tests, tonal memory is involved.

Using the California Test of Mental Maturity as a measure of general intelligence, we find little relationship between our tests and either the language or non-language parts of the California test. This leads to a substantiation of previous beliefs that only a low relationship exists between general intelligence and musical behavior.

This study has developed, we believe, a useful set of measures of musical behavior. Percentile ranks are available (see Appendix II) for either musicians or unselected freshmen, against which any individual score may be compared. These tests, used at the senior high school level or above, should be valuable in selecting the most qualified students for admission to music schools and colleges. For final proof of validity, however, a longitudinal study should be undertaken to determine the relationship between tests scores on entrance to music colleges and musical achievement during and after training.

APPENDIX I

Instructions to Subjects

Test I: Interval Discrimination

In music, an interval is the distance between any two notes or steps on the scale. The object of this test is to measure how well you can distinguish various intervals. In each item of this test, you will hear two intervals played in succession. You are asked to tell whether the two intervals are the same or different. When an interval is the same, the number of steps on the scale which lie between the first and second notes *and* the third and fourth notes will be the same. This does not mean that the actual *notes* will be the same, but the intervals are. If the second interval is the same as the first, encircle the letter *S* on your answer sheet opposite the appropriate number. If the second interval is different, encircle the letter *D*.

Here are some practice trials. Start with letter *A* at the top of your answer sheet under Test I. (*A* is played). This example was "different" because the second interval was larger than the first. (*B*—same; *C*—different.) In the first part of this test, the second note of each interval is higher than the first.

Part II. This is exactly the same as Part I with the exception that the second note of each interval is lower than the first. Here are some more practice exercises.

Test II: Melodic Transposition

When we say that a melody is transposed, we mean that it is played in a different key from the original; the actual notes are different but their relationships to each other are the same. In each item of this test you will hear two simple melodies. The second melody will always be in a different key from the first. Sometimes this transposed melody will be the same as the original; that is, the tonal relationship in each case is the same so that if the second melody were transposed back into its original key, the melodies would be the same in every respect. At other times the second melody will be different, that is, the relationship of the notes will be changed. If you think the second melody is the same as the first except for the change in key, encircle the letter *S* opposite the appropriate number on your answer sheet. If you feel the second melody is different from the first, encircle the letter *D*. In this test rhythmic patterns will always be the same, all changes which occur will be in the relationship of the notes. Here are some practice trials. Start with *A* in Test II (*A*—same; *B*—different; *C*—different).

Test III: Mode Discrimination

In music when notes are played together to form a chord, the particular combination of those notes determines its mode. You have probably heard of chords being spoken of as major or minor. Two chords are in the same mode, for example major, if their notes bear the same relationship to each other although they may be in different keys. In Test II, Melodic Transposition, notes were played separately, but items could be the same if the relationships of the notes were the same. In this test the only difference is that the notes are played together or simultaneously. In each item of this test you will hear two chords played. If you think both chords are the same, except for the fact that they are in different keys, encircle the letter *S* on your answer sheet. But if the two chords sound different, that is, they are in different modes the notes bearing a different relationship to each other, encircle the letter *D*. Here are some practice trials.

Test IV: Melodic Sequences

In each item of this test you will hear four separate groups of notes, each group following a similar melodic pattern. They are what musicians call sequences. Sometimes all groups will follow the *same* melodic pattern. Other-times, there will be a mistake or change in the fourth or last sequence. Listen carefully to the first three groups to get the pattern which is being followed. Then determine whether or not the fourth sequence follows the same pattern. If it does encircle the letter *S* in the appropriate place on your answer sheet indicating that all four sequences follow the same melodic pattern. If you detect a change or difference in the pattern of the last sequence, encircle the letter *D* indicating that its pattern was different. In this test the rhythm of each pattern will be the same. Here are some practice exercises.

Test V: Rhythmic Discrimination

This test is similar to the preceding one in that you will again hear sequences. However, here it is the rhythmic patterns which concern us. Each item will consist of four sequences each following a certain *rhythmic* pattern. In some cases the *4th* pattern will follow the same rhythmic grouping as the first three. In other cases it will change in rhythm. If you believe all four sequences have the same rhythm, encircle *S* in the appropriate place on your answer sheet. But if you note that the *last sequence* changes its rhythmic pattern encircle *D*. Listen to the first three sequences to see what rhythm is being followed and then determine whether the last sequence is the same or different. In all cases the melodic patterns are correct. Here are some practice exercises.

APPENDIX 2

Percentile Norms for Unselected and Musician Groups

Test I (total), Musicians: Interval Discrimination		Test I (total), Unselected: Interval Discrimination	
Score	Percentile Rank	Score	Percentile Rank
50	99	50	99
49	97	49	99
48	91	48	99
47	83	47	99
46	69	46	99
45	59	45	99
44	50	44	98
43	40	43	96
42	31	42	95
41	21	41	93
40	15	40	91
39	11	39	84
38	8	38	80
37	6	37	74
36	4	36	70
35	3	35	64
34	2	34	56
33	1	33	52
32	1	32	41
31	1	31	37
30	1	30	31
29	1	29	26
28	1	28	19
27	1	27	15
26	1	26	8
25	1	25	6
24	1	24	5
23	1	23	3
22	1	22	2
		21	1
		20	1

Test II, Musicians: Melodic Transposition		Test II, Unselected: Melodic Transposition	
Scores	Percentile Rank	Score	Percentile Rank
30	99	30	99
29	80	29	99
28	50	28	95
27	36	27	89
26	25	26	80
25	12	25	69
24	8	24	60
23	6	23	44
22	4	22	33
21	3	21	28
20	2	20	20
19	1	19	12
18	1	18	10
17	1	17	6
		16	4
		15	3
		14	1
		13	1

Test III, Musicians: Mode Discrimination		Test III, Unselected: Mode Discrimination	
Score	Percentile Rank	Score	Percentile Rank
30	99	30	99
29	97	29	99
28	94	28	99
27	91	27	99
26	84	26	99
25	76	25	99
24	71	24	98
23	62	23	97
22	51	22	96
21	37	21	95
20	26	20	92
19	21	19	85
18	14	18	77
17	9	17	66
16	7	16	52
15	6	15	37
14	5	14	21
13	3	13	11
12	1	12	3
11	1	11	1
10	1	10	1
9	1	9	1
8	1	8	1

Test IV, Musicians: Melodic Sequences		Test IV, Unselected: Melodic Sequences	
Scores	Percentile Rank	Scores	Percentile Rank
30	99	30	99
29	85	29	97
28	70	28	96
27	55	27	89
26	40	26	88
25	30	25	82
24	20	24	75
23	11	23	70
22	9	22	63
21	5	21	53
20	4	20	44
19	1	19	35
18	1	18	24
17	1	17	18
16	1	16	14
15	1	15	9
		14	5
		13	3
		12	1

Test V, Musicians: Rhythmic Sequences		Test V, Unselected: Rhythmic Sequences	
Scores	Percentile Rank	Scores	Percentile Rank
30	99	30	99
29	90	29	97
28	65	28	96
27	43	27	86
26	30	26	75
25	20	25	66
24	15	24	54
23	11	23	42
22	9	22	33
21	8	21	27
20	2	20	22
19	2	19	13
18	1	18	9
		17	8
		16	4
		15	3
		14	2
		13	1

Total Scores, Musicians:		Total Scores, Unselected:	
Score	Percentile Rank	Score	Percentile Rank
170	99	153–170	99
169	99	152	99
168	99	151	99
167	99	150	98
166	97	149	97
165	96	148	97
164	95	147	96
163	94	146	95
162	92	145	95
161	88	144	95
160	86	143	94
159	85	142	94
158	81	141	94
157	77	140	93
156	73	139	91
155	69	138	91
154	67	137	90
153	64	136	88
152	61	135	87
151	52	134	84
150	51	133	83
149	48	132	81
148	45	131	80
147	41	130	77
146	37	129	76
145	36	128	74
144	34	127	72
143	31	126	72
142	28	125	70
141	24	124	67
140	20	123	64
139	18	122	62
138	16	121	59
137	14	120	57
136	13	119	54
135	12	118	50
134	10	117	49
133	10	116	45
132	9	115	43
131	8	114	39
130	8	113	35
129	8	112	30
128	8	111	29
127	7	110	28
126	6	109	27
125	5	108	26
124	5	107	23
123	5	106	21
122	5	105	19
121	3	104	18
120	2	103	17
119	2	102	15
118	1	101	14
		100	12
		99	10
		98	9
		97	8
		96	6
		95	5
		94	5
		93	5
		92	4
		91	3
		90	3
		89	3
		88	2
		87	1

BIBLIOGRAPHY

1. Beinstock, S. F. A predictive study of musical achievement. *J. Genet. Psychol.*, 1942, **61**, 135-145.
2. Brennen, F. M. The relation between musical capacity and performance. *Psychol. Monogr.*, 1927, **36**, No. 167, 190-248.
3. Brown, A. W. The reliability and validity of the Seashore tests. *J. Appl. Psychol.*, 1928, **12**, 468-476.
4. Chadwick, J. E. Predicting success in sight-singing. *J. Appl. Psychol.*, 1933, 17, 671-674.
5. Drake, R. M. Four new tests of musical talent. *J. Appl. Psychol.*, 1933, 17, 136-147.
6. Drake, R. M. The validity and reliability of tests of musical talent. *J. Appl. Psychol.*, 1933, 17, 447-452.
7. Drake, R. M. *Drake test of musical talent.* Fredricksburg, Va., 1942.
8. Drake, R. M., and Farnsworth, F. R. A historical, critical and experimental study of the Seashore-Kwalwasser test battery. *Genet. Psychol. Monogr.*, 1931, **9**, No. 25, 291-391.
9. Drake, R. M., and Fransworth, F. R. Studies in the psychology of tone. *Genet. Psychol. Monogr.*, 1934, **15**, No. 1, 1-64.
10. Groves' *Dictionary of music and musicians.* New York: Macmillan Co., 1938.
11. Guilford, J. P. *Fundamental statistics in psychology and education.* New York: McGraw-Hill, 1942.
12. Highsmith, J. A. Selecting musical talent. *J. Appl. Psychol.*, 1929, **13**, 486-493.
13. Kwalwasser, J., and Dykema, P. W. *Manual of directions for Victor records.* Carl Fisher, 1930.
14. Lanier, L. H. Prediction of the reliability of mental tests and tests of special abilities. *J. Exp. Psychol.*, 1927, 10, 69-133.
15. Lundin, R. W. A preliminary report on some new tests of musical ability. *J. Appl. Psychol.*, 1944, **28**, 393-396.
16. Madison, T. H. *Interval discrimination as a measure of musical aptitude. Arch. Psychol.*, 1942, No. 206, 1-99.
17. Manzer, G. W., and Morowitz, S. The performance of a group of college students on the K-D tests. *J. Appl. Psychol.*, 1935, 19, 331-346.
18. McCarthy, D. A study of the Seashore measures of musical talent. *J. Appl. Psychol.*, 1930, **14**, 437-445.
19. Mosher, R. M. *A study of the group method of measurement of sight-singing.* Bureau of Publications, Teachers College, Columbia University. New York, 1925.
20. Mursell, J. L. Measuring musical ability and achievement: a study of the correlation of the Seashore test scores and other variables. *J. Educ. Res.*, 1932, **25**, 116-126.
21. Mursell, J. L. *The psychology of music.* New York: W. W. Norton and Co., 1937.
22. Mursell, J. L. What about music tests? *Music Educators J.*, 1937, **24**, No. 2, 16-18.
23. More, G. V. D. Prognostic testing in music on the college level. *J. Educ. Res.*, 1932, **26**, 199-212.
24. Saetveit, J. C., Lewis, D., and Seashore, C. E. *Revision of the Seashore measures of musical talent.* Univ. Iowa Studies: Aims and progress of research, 1940, No. 65, 1-66.
25. Salisbury, F. S., and Smith, H. R. Prognosis of sight-singing ability of normal school students. *J. Appl. Psychol.*, 1929, **13**, 425-439.
26. Sanderson, H. E. Differences in music ability in children of different national and racial origin. *J. Genet. Psychol.*, 1933, **42**, 100-120.
27. Schoen, M. Tests of musical feeling and understanding. *J. Comp. Psychol.*, 1925, 5, 31-52.
28. Seashore, C. E. *Manual of instructions and interpretations for measures of musical talent.* Chicago: Stoelting Co., 1919.
29. Seashore, C. E. *The psychology of musical talent.* Newark: Silver, Burdett, 1919.
30. Seashore, C. E. *The psychology of music.* New York: McGraw-Hill 1938.
31. Seashore, C. E. Psychology of music, XXI. Revision of the Seashore measures of musical talent. *Music Educators J.*, 1939, **26**, 31-33.
32. Stanton, H. M. *Measurement of musical talent: The Eastman experiment.* Univ. Iowa Stud. Psychol. Music, 1935, **2**, 1-140.
33. Taylor, E. M. A study of the prognosis of musical talent. *J. Exp. Educ.*, 1941, 10, 1-28.
34. Tilson, L. M. Music talent tests for teacher-training purposes. *Music Superv. J.* 1932, **18**, 26.
35. Whitley, M. L. A comparison of the Seashore and K-D tests. *Teach. Coll. Rec.*, 1932, **8**, 731-751.
36. Wilson, M. E. The prognostic value of music success of several types of tests. *Music. Superv. J.*, 1930, **16**, 1-83.
37. Wright, T. A. The correlation between achievement and capacity in music. *J. Educ. Res.*, 1929, **17**, 50-56.